DOCUMENT RESUME

ED 210 073 JC 810 654

AUTHOR Rasor, Richard A.; And Others
TITLE Student Evaluations of Their Best and Worst College Experience.
INSTITUTION American River Junior Coll., Sacramento, Calif.
PUB DATE Jun 81
NOTE 25p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; Community Colleges; *Evaluation Criteria; Evaluation Methods; *Questioning Techniques; *Questionnaires; *Student Evaluation of Teacher Performance; Two Year Colleges; Two Year College Students

ABSTRACT

A study was conducted at American River College to design a teacher evaluation instrument that would focus on students by asking them to describe what happens to them in the classroom and would restrict students' evaluations of their instructor to observable behavior. Over 300 potential questionnaire items were generated from a review of the literature, from other teacher evaluation forms, and from the responses of over 150 students who were asked to identify three characteristics of their best and worst teachers. From this item pool, 44 items were selected based on the following criteria: (1) the items could be reworded to describe students' in-class learning, behaviors, and feelings; (2) the items were appropriate for a frequency of occurence or yes/no response scale; and (3) the items avoided inferring instructor traits. A total of 213 students were asked to complete evaluations of their best and worst teachers by stating, for example, the frequency that they felt they were "able to take meaningful notes in class" or their degree of agreement with statements such as "I don't see how this subject is important to my life." Correlation and multiple regression analyses were conducted to determine the relationship between each questionnaire item and the "best" and "worst" conditions. On the basis of these analyses 22 questionnaire items were recommended as reliable for teacher evaluation forms. The study report reviews arguments concerning teacher evaluations, presents statistical findings, and includes the recommended instrument. (AYC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED210073

STUDENT EVALUATIONS OF THEIR BEST AND WORST COLLEGE EXPERIENCE

by

Richard A. Rasor

Kim Lundquist

Dave Miller

June 1981

JC 810 654

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard A. Rasor

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Over the years there has been a substantial number of studies done regarding college teacher evaluation. Undoubtedly there are many opinions over what constitutes effective teaching and there probably is no mutually agreed upon answer. Still, faculty have agreed upon some desirable outcomes for their students. The real differences of opinion center around how to measure good teaching and who should make the judgment. Most measuring attempts have utilized student evaluations of teacher behavior. Hundreds of questionnaire items have been written, subjected to thorough statistical analysis and put into objective survey instruments that students fill out in an attempt to assess the effectiveness of teaching. There is little question that such instruments are highly reliable (Costin, Greenough and Menges, 1971; French-Lazovik, 1974). But what about validity? Tomasco (1980) correlated student ratings with personality variables of teachers. It was found that teachers with favorable evaluations have specific personality characteristics which students consistently identify. Tomasco states:

> "For example, it seems reasonable to suggest that if a teacher shows a great deal of respect for student opinions, manifests sensitivity to student problems and concerns and reflects a genuine interest in the progress of his class, then students are likely to perceive that teacher as friendly, helpful and willing to put forth considerable effort in maintaining the well being of that class; likewise, it would seem plausible to assume that teachers who effectively challenge their students and who perform their daily teaching tasks with precision, efficiency, and organization would also be seen by their students as productive, achievement-oriented, perfectionistic and constructively agressive." (p.81)

Marsh and Overall (1980) examined the validity of student ratings with respect to multiple criteria both at midterm time and at the end of the term. They concluded that student evaluations are valid, but midterm ratings while useful as feedback to instructors, are consistently less correlated with teaching effectiveness criteria than ratings collected at the end of the course.

Lamberth and Kosteski (1981) recognized the problem of using students whose ratings may be influenced by their own classroom achievment. The argument suggests that students who receive good grades may give the teacher a higher evaluation than students receiving poorer grades regardless of the teaching skills exhibited by the instructor. In fact, Shingles (1977) has proposed doing away with all student evaluations if this variable cannot be controlled in the sense of adjusting the ratings accordingly. Student achievement and grades may be conceptualized as a source of bias that could invalidate teacher evaluations or as one criterion by which to judge teacher effectiveness. Lamberth and Kostelski (1981) compared student evaluations of teachers with ratings also given by teaching assistants. The TAs were doctorate candidates who attended every lecture but who were supervised by an independent professor. Hence the TAs were not subject to the pressures to evaluate highly because of grades, they were knowledgeable about the subject matter being taught and had extensive in-class contact with the teachers being evaluated. The authors concluded that ratings by TAs and students are highly correlated and stable across two independent sets of students. This gives strong argument for the validity of student evaluations.

All the validation efforts are laudable but one fact remains: faculty have certain misgivings about students judging their competency -- especially if such evaluations become public information or are used in decisions regarding tenure and promotion. Cries of "popularity contest" capture the protest flavor of those who object to student raters. Yet there is little doubt that such ratings are both reliable and reasonably valid. Probably the objections to student ratings will always be present regardless of their usefulness. Any evaluation process is likely to be personally threatening when adverse consequences could be a result. Furthermore, students are not in a position

to judge the instructor's knowledge of subject matter, and some of the evaluation items are overly vague, e.g., The instructor "challenges students" or "inspires confidence." Most of the objections against student evaluations could be lessened somewhat if students were asked only to describe what is happening to them in the classroom and restrict their perceptions about the instructor to that which is observable. Instead of asking "Are the instructor's lectures interesting?"; why not ask "How often do you have difficulty keeping your mind on what is happening in class?" The main difference in the two types of questions is that one asks the student to evaluate while the other asks for facts about student performance. Hayden (1979) put together a 21-item questionnaire following this approach. He reports that students seem relieved in being asked only for their factual experience. It is also important that teachers know what students are experiencing in class. Unfortunately, Hayden (1979) does not provide any research data for his questionnaire so one cannot be sure that his items are the best.

We believe that student self-description is the correct teacher evaluation path to follow yet data must be collected to initailly start the process of reliability and validity. The purpose of our research was to design an instrument which shifts the evaluation focus into first person (the student rater) and restricts those perceptions of instructors to observable behavior.

Method

Subjects

A total of 213 students enrolled in several lower division classes at American River College during the Spring, 1981 semester served as subjects. Nearly all of them were given a small point bonus towards their final grade as a reward for filling out the questionnaire on their own time. The sample

consisted of 78 males and 135 females with a median age of 21.74 years. Collectively the median number of college units completed was 21.68 with 24% having completed 45 units or more. Only 12% were in their first semester of college. It is reasonable to conclude that the sample had an adequate amount of community college experience to take part in this type of research.

Questionnaire Construction

Items were generated from a variety of sources; from over 150 students who wrote down three characteristics of their "best ever" college teacher as well as their "worst ever," from numerous research articles e.g., French-Lazovik (1974), and from a variety of existing questionnaires used in teacher evaluation. Guidelines for final item selection were as follows: (1) The item should be reworded to describe the student's cognition, feeling or behavior that is exhibited in class (2) The item must be appropriate for a frequency of occurrence or a yes-no response scale (3) The item should avoid inferring presence of a trait in the instructor, e.g. "Inspires leadership or "Well prepared" (4) The item should concentrate upon observable teacher behavior, e.g., "Hands out assignments at the beginning of class" (5) Feedback on the item must be of value to the instructor. From a pool of over 300 possible items, 44 were sufficiently developed to be of use in this study. Half of the items were written favorably toward the learning situation while the others were worded unfavorably. This was done to minimize response bias, and some items are more informative in the negative, e.g., "I am confused during lectures."

Procedure

The students were asked to fill out the questionnaire by rating each item for their "best ever" college teacher. Following that they rated their "worst ever" teacher with the same items only numbered differently. Responses on the answer sheets where then keypunched for computer analysis.

Results and Discussion

Means, standard deviations for each item under both best and worst instructor conditions appear in Table 1. In addition, rank order of mean differences (1 = largest difference between means) and simple correlation of each item with criterion (best versus worst) appear in Table 1. All 44 items correlated well with the forced criterion (Median = .57) with a range of .32 to .78. Under the best condition there was less dispersion on the items in that the standard deviations are smaller than the same item under the worst condition. This was true in all but two instances. Also, in 30 of 44 items, over 80% of the best ratings were in two categories of the 5-point scale. This would suggest that student raters are in more agreement over their best instructor condition than under the worst situation. Whether this reflects bias in the instrument, halo effect, or lack of a true worst instructor experience cannot be answered. In Table 2 are presented multiple correlation data. The highest four items alone accounted for 72% of the variance. Thus, balance of breadth and detail, good illustrations and examples given in class, students' feeling of learning something, and prompt feedback on tests and assignments just about explain all the variance. Other items certainly correlate with the criterion, but they also highly intercorrelate with the four items mentioned, hence they appear to be measuring the same thing -- best instruction. Because of the high intercorrelations, remaining items add little to predictive power. The full model (all 44 items) resulted in an R^2 of .79. Estimated shrinkage on another sample reduces this proportion of variance explained to 74% but cross-validation is needed for accurate appraisal.

The grades received with the best instructor were: A = 58%, B = 36%, C = 5%, D, F, NC, INC, drops = 1%. With the worst instructor grades were quite different: A = 17%, B = 24%, C = 25%, D = 12%, F or NC = 9%, INC = 1% and

drops = 13%. Grades accounted for 26% of the variance in best vs worst classification. This could be a source of bias in student evaluation as Shingles (1977) has suggested, or it may reflect the notion that students work harder and achieve higher levels of performance with instructors who inspire them. Marsh (1977) did not find a significant student grade differential between best versus worst instructors, but did find a substantial agreement between college graduating seniors' nominations for best and worst instructors and subsequent classroom evaluations using different student raters. He concluded that student evaluations are valid and that background variables (e.g., course workload, course difficulty, interest in subject, expected grade) play an insignificant role in differentiating the best from the worst instructors.

In Table 3 are the items we would recommend for initial use in the classroom. Naturally validation and reliability of this shorter version needs to be established. Also, two written response forms to accompany the questionnaire are presented. The final selection of items for recommended use proceeded along empirical guidelines. Eighteen of the twenty items had the highest correlations with the criterion, all items that contributed to prediction in multiple regression were used, and all had a significant difference in means between best versus worst conditions. Two new dependent variable items were also added. The recommended items also fit well into broad categories that have been established through detailed factor analysis: enthusiasm and concern of teacher, breadth of coverage, organization of lectures, interaction with student, and students' value derived from learning (Marsh and Overall 1980).

The final questionnaire with accompanying written forms for students and instructor should provide a meaningful feedback exchange between the two parties. Hopefully such communication will result in improved learning.

TABLE 1

Item Statistics for 44 Statements in Rank Order of Correlation Magnitude with Criterion[1]

Item[2]	"Best" Condition		"Worst" Condition		Rank Order Absolute Mean Difference	Pearson Correlation with Criterion
	Mean	SD	Mean	SD		
There is a good balance of breadth & detail in this class	4.29	.77	2.23	1.04	3	.73
The examples & illustrations in lecture help make difficult concepts clear to me	4.37	.74	2.35	1.09	4	.75
I feel that I'm learning something and not just going for a good grade	4.67	.61	2.41	1.31	1	.75
I believe the sequence of reading assignments and lectures is good	4.34	.74	2 37	1.17	5	.71
I like the instructor's sense of humor	4.44	.82	2.34	1.18	2	.71
I have sense of direction about how the lectures and readings will lead me rrom one level of understanding to another	4.16	.79	2.37	1.11	11.5	.70
By attending class I learn additional information and not just a repeat of the textbook	4.48	.66	2.59	1.22	6	.69

[1]Items 1-26 used a frequency scale where 1 = almost never, 2 = seldom, 3 = occassionally, 4 = frequently, 5 = almost always
Items 27-44 used an agreement scale where 1 = definately no, 2 = most likely no, 3 = not sure, 4 = most likely yes, 5 = definitely yes

[2]All items with an asterisk are worded unfavorably so mean values under "worst" condition should be higher

TABLE 1 (Continued)

Item	"Best" Condition		"Worst" Condition		Rank Order Absolute Mean Difference	Pearson Correlation With Criterio
	Mean	SD	Mean	SD		
The lecture pace seems right for me	4.34	.76	2.46	1.04	7	.78
I feel respected as an individual in this class	4.44	.78	2.64	1.16	9	.69
I have trouble paying attention in class*	1.82	.95	3.61	1.15	11.5	-.67
I receive encouragement from the instructor's advice or praise	3.58	1.11	1.85	1.09	14	.65
I have a clear understanding of what I need to study and know for tests	4.47	.81	2.68	1.29	11.5	.64
I dislike the lecture content*	1.78	.90	3.46	1.19	15	-.63
I have trouble getting help from the instructor*	1.56	.88	3.35	1.30	11.5	-.63
I feel there is too much wasted time in class*	1.62	.83	3.45	1.29	8	-.62
I get prompt feedback from tests and assignments	4.32	.83	2.70	1.28	17	.61
I am unable to see how the material interrelates*	1.66	.94	3.11	1.09	22	-.60
My reading is coordinated with the lecture topics	4.37	.81	2.84	1.38	19.5	.59

TABLE 1 (Continued)

Item	"Best" Condition		"Worst" Condition		Rank Order Absolute Mean Difference	Pearson Correlation With Criterion
	Mean	SD	Mean	SD		
I feel that I cannot express criticism or disagreement to the instructor*	1.75	.97	3.38	1.30	16	-.59
I am discouraged to be creative in this class*	1.69	.98	3.06	1.15	24.5	-.59
I believe the instructor is aware of my progress even without consulting the record	3.81	1.03	2.20	1.21	18	.59
The material written on the chalkboard (or projected onto the screen) is unclear to me*	1.71	.87	3.05	1.16	26.5	-.57
I need more explanation of some concepts presented in class in order to understand them thoroughly*	2.38	.88	3.72	1.13	26.5	-.56
I feel (or sense other students feel) "put down" or insulted in class*	1.50	.86	3.01	1.29	21	-.55
I enjoy reading the textbook (or assigned readings)	3.65	1.03	2.28	1.13	24.5	.54
I am able to take meaningful notes from the lecture	4.10	1.11	2.57	1.25	19.5	.54
I enjoy the visual media used in class (e.g., films, slides, videotapes)	4.08	1.11	2.69	1.27	23	.51
I feel comfortable asking questions or making comments in class when the need arises	3.83	1.29	2.53	1.29	29	.49

TABLE 1 (Continued)

Item	"Best" Condition		"Worst" Condition		Rank Order Absolute Mean Difference	Pearson Correlation With Criterion
	Mean	SD	Mean	SD		
I have difficulty seeing how the lecture material is related to the theme of this course*	1.70	.96	2.84	1.24	34	-.49
I don't see how this subject is important to my life*	1.56	.97	2.87	1.41	28	-.48
I have little understanding of what the requirements are to earn a specific final grade*	1.66	1.07	2.93	1.33	31	-.48
The required reading is beyond my clear understanding*	1.72	.81	2.81	1.22	35	-.46
I believe the final grade is dependent upon too few assignments or tests*	1.82	1.08	3.11	1.38	30	-.46
I believe there is unwarranted criticism of other subject areas or other faculty made in class*	1.46	.81	2.33	1.27	43	-.45
I feel uncomfortable about talking informally with the instructor*	2.17	1.27	3.43	1.36	32	-.45
I have problems just understanding the test questions*	1.95	.89	2.97	1.23	38	-.45
The class atmosphere makes it difficult to get to know other students*	1.90	1.04	3.05	1.34	33	-.44
I believe the instructor has reasonably high expectations of me	3.69	.92	2.71	1.19	39	.43

TABLE 1 (Continued)

Item	"Best" Condition		"Worst" Condition		Rank Order Absolute Mean Difference	Pearson Correlation With Criterion
	Mean	SD	Mean	SD		
For my own enjoyment, I read more about this subject than is required	2.78	1.07	1.18	1.07	40	.42
The workload for earning the grade I want seems unreasonable*	1.91	1.12	2.96	1.32	36.5	-.39
I make known to the instructor complications or conflicts I have with this class	3.54	1.22	2.59	1.27	41	.36
I have enough time to complete the tests	3.59	.83	3.65	1.33	42	.36
I am unsure about the due dates for assignments*	1.78	1.21	2.83	1.35	36.5	-.35
A few students are allowed to dominate discussion*	2.21	.97	2.93	1.29	44	-.32

TABLE 2

Multiple Correlation Summary

Item	Multiple R	R-Square	R-Square change
There is a good balance of breadth & detail in this class	.777	.604	.604
The examples and illustrations in lecture help make difficult concepts clear to me	.820	.672	.068
I feel that I'm learning something and not just going for a "good" grade	.837	.700	.028
I get prompt feedback from tests and assignments	.849	.720	.020
I feel there is too much wasted time in class	.856	.733	.013
I receive encouragement from the instructor's advice or praise	.863	.744	.011
I need more explanation of some concepts presented in class in order to understand them more thoroughly	.867	.752	.008
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
All items	.890	.792	.00002

TABLE 3

Recommended Items for Instructor-Student Feedback Questionnaire

A or 1 = (almost) Never
B or 2 = Seldom
C or 3 = Occasionally
D or 4 = Frequently
E or 5 = (almost) Always
Leave blank if item does not apply

1. By attenting class I learn additional information and not just a repeat of the textbook.
2. The lecture pace seems right for me
3. I have trouble paying attention in class
4. I an unable to see how the material interrelates
5. I like the instructor's sense of humor
6. I receive encouragement from the instructor's advice or praise
7. I dislike the lecture content
8. I have trouble getting help from the instructor
9. I get prompt feedback from tests and assignments
10. I am able to take meaningful notes from the lecture
11. I feel there is too much wasted time in class
12. I need more explanation of some concepts presented in class in order to understand them thoroughly

New Scale

A or 1 = Definitely no
B or 2 = Most likely no
C or 3 = Not sure
D or 4 = Most likely yes
E or 5 = Definitely yes
Leave blank if item does not apply

13. There is a good balance of breadth and detail in this class
14. The examples and illustrations in lecture help make difficult concepts clear to me
15. I feel that I'm learning something and not just going for a "good" grade
16. I feel that I cannot express criticism or disagreement to the instructor

TABLE 3 (Continued)

A or 1 = Definitely no
B or 2 = Most likely no
C or 3 = Not sure
D or 4 = Most likely yes
E or 5 = Definitely yes
Leave blank if item does not apply

17. I believe the sequence of reading assignments and lectures is good

18. I have a sense of direction about how the lectures and readings will lead me from one level of understanding to another

19. I feel respected as an individual in this class

20. I have a clear understanding of what I need to study and know for tests

Summary items

21. Assume a close friend needs to take this same course, wants a meaningful educational experience, but has to select one instructor from several. Would you recommend your current instructor?

22. Overall, would you rate this entire course a good experience?

Please continue with the sheet asking for your opinions.

Thank you

TABLE 3 (Continued)

YOUR OPINIONS PLEASE!

From your responses to the items on the questionnaire, please elaborate (give suggestions, comments, insights, recollections of critical instances) on 3 items you rated the highest and 3 items you rated the lowest. Please refer to items by number.

Also, additional comments would be appreciated on areas of instruction not covered by the questionnaire. Thank you.

HIGHEST RATED ITEMS (major strengths of the class)

Item#

Item#

Item#

LOWEST RATES ITEMS (most room for improvement)

Item#

Item#

Item#

ADDITIONAL SUGGESTIONS FOR IMPROVING COURSE.

TABLE 3 (Continued)

Feedback to Students

Comments from your Instructor:

The highest rated items concerned:

The lowest rated items concerned:

My response to the above is:

References

Costin, F., Greenough, W. T. & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

French-Lazovik, G. Predictability of students' evaluations of college teachers from component ratings. Journal of Educational Psychology, 1974, 66, 373-385.

Hayden, J. O. Teacher evaluation: a very different approach. Teaching Resources Center monograph 8, University of California - Davis, 1979.

Lamberth, J. & Kosteski, D. M. Student evaluations: an assessment of validity. Teaching of Psychology, 1981, 8, 8-11.

Marsh, H. W. The validity of students' evaluations: Classroom evaluation of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 1977, 14 (4), 441-447.

Marsh, H. W., & Overall, J. U. Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. Journal of Educational Psychology, 1980, 72, (4), 468-475.

Shingles, R. D. Faculty ratings: Procedures for interpreting student evaluations. American Educational Research Journal, 1977, 14 (4), 459-470.

Tomasco, A. T. Student perceptions of instructional and personality characteristics of faculty: A canonical analysis. Teaching of Psychology, 1980 7 (2), 79-86.

DEC 2 3 1981

ERIC Clearinghouse for Junior Colleges
96 Powell Library Building
University of California
Los Angeles, California 90024